基于 Rao-Stirling 指数的学科交叉文献发现*
——以纳米科学与纳米技术为例

韩正琪[1,2] 刘小平[2] 徐涵[1,2]
[1] 中国科学院大学 北京 100049 [2] 中国科学院文献情报中心 北京 100190

摘要：［目的/意义］学科交叉文献发现是进行学科交叉研究的重要前提，从海量的文献中快速、精准地发现领域相关交叉文献有助于研究人员快速地把握领域学科交叉动态，识别领域学科交叉研究热点与前沿。提出基于 Rao-Stirling 指数的领域学科交叉文献发现方法，并以纳米科学与纳米技术领域为例，探讨该方法的可行性。［方法/过程］在 Web of Science 数据库下载纳米科学与纳米技术领域文献，构建期刊缩写－全称－学科类别对照表，利用 Python 编程构建文献参考文献学科分布矩阵，利用 R 编程计算每篇文献的 Rao-Stirling 指数进行文献的学科交叉测度，根据测度结果将纳米科学与纳米技术领域文献按照学科交叉程度分为三个水平，以发现领域学科交叉文献。［结果/结论］基于 Rao-Stirling 指数的领域学科交叉文献发现方法可以实现领域文献水平的学科交叉测度，并发现学科交叉文献，且该研究方法也同样可扩展到其他研究领域。

关键词：学科交叉 Rao-Stirling 学科交叉文献 纳米科学与纳米技术
分类号：G250
DOI：10.13266/j.issn.0252-3116.2018.01.016

1 引言

随着科学研究的不断深入，学科不断分化，学科逐渐细分为许多子学科，向着更加专业化的方向发展。然而在继续分化的同时，学科又逐步走向交叉综合，向着高度综合化、整体化、交叉化的方向发展。吸收不同学科领域的知识被认为是科学发展的关键因素。面对重大社会问题和复杂的科学问题，往往需要借助几个学科之间的交叉、渗透、融合来共同解决。

学科交叉研究（interdisciplinary research，IDR），又称为跨学科研究或者交叉科研，即通过学科间知识的整合而进行的研究[1]。学科交叉研究通常意味着创新和进步，学科交叉点往往是科学新的研究热点和前沿。物理学家海森伯说道："最突出的科学发现往往发生于两条不同的思维路线的交叉点上"[2]。学科交叉研究可以解决各传统学科所不能触及的未知领域的研究问题，在传统学科已经发展成熟的情况下，这些未知领域往往是重要学术突破的产生之处。学科交叉研究是完成复杂高难科研问题的有效选择，DNA 双螺旋结构、基因重组技术、核磁共振成像等许多重要的科学发现都是因为打破了学科界限而获得。科学最高奖诺贝尔奖 1900－2011 年颁发的 365 项自然科学奖项中，有 198 项为学科交叉领域，所占比例高达 54.2%，并且呈稳步上升趋势[3]。每年《科学》（*Science*）杂志评出的"十大科学突破"中绝大多数也为学科交叉研究的成果。

随着学科交叉研究在科技进步中扮演越来越重要的角色，从不同的角度来研究学科交叉是非常必要的。对某领域相关的学科交叉点进行识别，揭示学科交叉关系和特点，把握学科交叉研究的热点方向和趋势，已成为学科交叉研究的重要分支，可以为相关研究者提供理论思路和假设，帮助其选择正确的研究方向。并对高校的学科设置与调整和学科交叉研究中心的设立提供支持，为科学管理者项目评审、政府机构支持政策的制定提供借鉴。

2 国内外相关研究现状

对学科交叉进行的研究主要包括理论研究和实证

* 本文系中国科学院文献情报能力建设专项子课题"重大基础交叉前沿战略情报研究服务体系"（项目编号：Y170421001）研究成果之一。
作者简介：韩正琪（ORCID：0000－0002－7998－9717），硕士研究生；刘小平（ORCID：0000－0002－3342－8041），副研究员，硕士生导师，通讯作者，E-mail：liuxp@mail.las.ac.cn；徐涵（ORCID：0000－0002－0896－1637），硕士研究生。
收稿日期：2017－07－05 修回日期：2017－10－19 本文起止页码：125－131 本文责任编辑：王善军

研究两部分。理论研究主要为探讨学科交叉起源、演化、形成以及发展机制等,实证研究则主要基于文献计量或文本挖掘的方法发现某一期刊、机构或者学科的学科交叉情况,本文主要探讨实证研究部分。按照研究对象的不同,可以将学科交叉的文献发现研究分为四类,分别是基于关键词、基于文献自身、基于作者以及基于引文的学科交叉研究。

基于关键词的学科交叉研究将论文关键词作为共现单元,对关键词间的共现频次进行统计,共现频次越高代表关键词所属的学科领域间关联强度越高,领域间学科交叉性越强。如 A. Rip 等[4]以十年内生物技术领域核心期刊刊载的文献为研究对象,通过共词分析研究领域的动态变化,探索不同研究领域间联系随时间的变化。魏建香[5]利用 CNKI 数据库中近十年情报学和计算机科学两个领域核心期刊中的文献题录信息进行挖掘,借助模糊 C－均值聚类(FCM)算法发现两领域内的学科交叉文献,并利用基于免疫选择的粒子群优化(IS-PSO)算法与 FCM 的聚类融合算法对学科交叉文献进行知识发现和可视化展示。李长玲等[6]基于时间序列分析的方法,通过对时序高频词聚类,动态的展示情报学和计算机学两个学科交叉研究主题随时间的变化。基于关键词的学科交叉研究通常使用文本聚类的方法进行学科交叉研究,维度越多,算法复杂度越高,很难适用于较大规模数据的学科交叉研究,同时也难以识别词间语义关系。

基于文献自身的学科交叉研究以共类分析为基础,根据已有的分类标准(杜威十进制分类、Web of Science 学科类别等)将文献或期刊划分到相应的类别中。通过统计分析文献或期刊所属的不同学科类别来反应这些论文或期刊的学科交叉特性。如吴蕾等[7]以 Web of Science 中文献的学科分类作为判断学科交叉文献的依据,选取学科分类为农学与遗传学文献作为学科交叉文献进行主题发现与演化分析。A. Allan 等[8]基于研究者的研究输出(书籍、期刊、研究报告等)的杜威十进制分类来判定中西部州立大学研究活动的学科交叉程度。R. J. W. Tijssen[9]基于共类分析探讨能源领域学科交叉水平以及领域间学科交叉联系强弱,并进行可视化展示。基于文献自身的学科交叉研究测度方法较为简单,适合较大范围,多领域的学科交叉测度。但是由于数据库管理的时滞性,不能用于最近的研究。此外,共类分析基于已有的分类标准将文献划分为不同的学科,对于学科边界较为明显的领域,其研究结果较好,而对于学科边界比较模糊的文献,往往不能进行准确的划分,导致分析结果不准确。

基于作者的学科交叉研究即通过作者跨学科发文数或者文献内合作者所属不同学科数来评估学者、文献、期刊以及领域的学科交叉研究情况。J. Qin 等[1]通过作者、作者机构以及所属学科的数量以及合作的类型对研究合作进行分析,探讨不同水平和学科的作者合作研究的学科交叉程度的差异性。J. Schummer[10]通过合著者所属机构院系来确定研究者所属学科,并筛选合著者所属不同学科的文献作为学科交叉文献,研究探讨纳米科学和纳米技术领域学科交叉研究模式和程度,并调查不同机构、地理空间的研究合作情况,提出学科交叉领域与经典学科研究之间的差异和联系。基于作者的学科交叉研究虽能够弥补分类标准的局限性,并且能够揭示更多的学科交叉研究的综合认知和社会现象,但是由于将作者划分为某一学科依赖专家判断和大量的人工繁琐工作,只适用于小样本测度研究,大大降低了指标的实用性。

基于引文分析测度学科交叉的多样性是最常用的学科交叉测度方法,其认为一篇文章的引文所属的不同学科类别数可以表示该文章的学科交叉程度,而文章引文所属的学科可以通过刊载该文章期刊的学科类别表示。D. E. Chubin 等[11]通过引文类别分析区分学科交叉研究文献与传统单一学科研究文献,并且探讨学科交叉研究的特征,构建敏感的学科交叉测度指标。A. L. Porter 等[12]选择《期刊引证报告》中 19 个期刊的 383 篇文献进行研究,提出外类别引用法这一新的指标对人口统计学、运筹学/管理学和毒理学三个学科类别进行学科交叉测度,发现引文涵盖广泛学科领域的文献是极其少见的。J. M. Choi 等[13]对过去 20 年内美国天文学领域核心期刊中的文献进行引文分析,探讨学科内部以及学科交叉交流模式。魏海燕等[14]统计情报学领域文章及其引文的学科分布,采用学科影响度指标,研究与情报学研究相关的学科及其随时间的变化。研究人员通常会以引用参考文献的形式说明其研究受何种学科知识的启发,基于引文进行文献学科交叉测度能够从根本上识别学科交叉文献,更好的把握学科交叉发展历程。

目前有关学科交叉的实证研究或探讨文献或期刊层面的学科交叉特性,或研究机构层面的跨学科合作情况,而对于某一领域或学科层面的学科交叉研究,相关研究仅有从宏观上把握领域或学科与其他学科的交叉情况[15]或者对某两个领域或学科进行学科交叉的主题识别[4],而尚未有某领域相关的所有学科交叉主

题的识别研究。

研究某领域学科交叉的主题，把握某领域相关的学科交叉点，首先需要确定该领域的学科交叉文献。相关研究如郭凤娇等[16]利用情报学和计算机科学领域内核心期刊刊载论文中的关键词构造向量空间模型，通过模糊 C 均值算法（FCM 算法），发现情报学和计算机科学领域的学科交叉文献。而该方法仅能识别某两学科间的学科交叉文献，而无法识别出与某领域相关的所有学科交叉文献。如何从海量文献中发现某领域的全部学科交叉文献，并筛选学科交叉程度较高的文献作为进一步学科交叉主题识别的研究对象为研究的关键点。

3 研究方法

本研究基于引文分析进行某领域学科交叉的文献发现研究。基于引文分析测度研究领域学科交叉多样性的研究所使用的测度指标通常来自其他研究领域，特别是生态学中的生物多样性测度，主要的测度方法如表 1 所示。目前所使用的指标大多为第一代生物多样性测度指标 Gini-Simpson 指数和香农熵[17]以及第二代生物多样性测度指标 Rao-Stirling 指数[18]。

表 1 基于引文分析的交叉学科研究的常用指标

指标	定义或描述
多样性	文章引用的不同学科的数量
熵	量化参考文献的学科类别多样性程度
Gini-Simpson 指数	测度参考文献是怎样分布（或者集中）在学科类别中的
Rao-Stirling 指数	可被理解为考虑距离/相似度（学科间）的 Simpson 多样性

第一代生物多样性测度指标又被称为分布敏感性测度指标。最简单的测度指标为多样性，即测度文章应用不同学科的数量，认为一篇具有参考文献所属两个以上学科类别的文献比参考文献仅属于一个学科类别的文献更具有多样性，也就是学科交叉程度越高，这个方法的一个明显的局限性是它忽略了引文所属学科类别之间的分布情况。第一代多样性测量方法已被发展为考虑文献所属不同类别参考文献的数量，两个主要的指标为熵和 Gini-Simpson 指数，分别测度文献参考文献所属学科类别的不确定性和属于不同学科类别的概率。虽然第一代生物多样性测度指标考虑参考文献的学科分布，但是由于其基于所有学科类别是完全不同的这一假设而被批判。其认为不管文献参考文献所属学科类别之间的相似性如何，如果两篇文献的参考文献有着相同数量的学科类别，那么这两篇文献的学科交叉程度相同。

第二代生物多样性测度指标又被称为分布和相似性敏感的测度指标，其认为如果一篇文献参考文献所属学科类别密切相关，那么该文献的学科交叉程度较低；而如果一篇文献的参考文献所属学科类别非常不同，那么该文献为高度学科交叉文献。主要测度指标为 Andy Stirling 提出的 Rao-Stirling 指数，考虑学科间的相似性和相异性。相关公式如下：

$Variety = v_{max}$ = 文章参考文献所属的 WOS 学科类别数

$$Disparity = \frac{1}{v(v-1)} \sum_{i,j} d_{ij}$$

$$Balance = \frac{1}{\ln(v)} \sum_{i} p_i ln p_i$$

$$Rao\text{-}Stirling\ Diversity = \sum_{i,j} p_i p_j d_{ij}$$

其中，v_{max} 为数据集中文献参考文献所属学科类别最多的个数，p_i 为第 i 学科类别在参考文献中所占比例，d_{ij} 为学科类别 i 和学科类别 j 之间的距离。Variety（种类）表示文献的参考文献可以被归类到学科类别的数量，学科类别数量越多，则该文献的学科交叉性越强；Balance（平衡）描述参考文献的学科分布的均衡度，在这里使用 Shannon 均匀度表示，如果一篇文献中每个类别中参考文献个数是相同的，说明该文献的参考文献的学科分布是均衡的，学科分布越均衡，文献的学科交叉性越强；Disparity（差异）表示元素分布的特异性程度，说明文献引用不同学科分类的引文的差异程度（距离），如果一篇文献参考文献所属学科类别的相似度较大（例如物理化学和应用物理），那么该文献的差异度（Disparity）就较小，文献学科交叉程度就越小，如果相似度较小（例如动物学和应用物理），该文献的差异度较大，文献学科交叉程度就越大。前三个公式分别代表学科多样性的三个不同方面，Rao-Stirling 指数整合三个学科交叉度量层面，将多样性的种类（Variety）、平衡（Balance）和差异（Disparity）这三个维度结合在一起。

本研究拟以纳米科学与纳米技术领域为例，基于 Rao-Stirling 指数计算该领域文献引文所属期刊的学科分布来测度文献的学科交叉程度，从而探讨纳米科学与纳米技术领域文献的学科交叉情况并发现该领域的学科交叉文献。

4 数据来源与处理

4.1 样本数据来源

本文选择汤森路透 Web of Science（WoS）的 SCI-

EXPANDED 数据库中有关纳米科学与纳米技术的文献作为研究对象,检索式为 WC = (“Nanoscience & Nanotechnology”),并限定文献类型为 ARTICLE、PROCEEDINGS PAPER 和 REVIEW,出版年为 2016 年,检索日期为2017 年3 月31 日,检索结果为40 618。利用 L. Leydesdorff 开发的 WoS 文献题录信息处理程序 ISI.exe 提取下载文献的题录信息,并生成关系数据库[19]。

4.2 构建期刊缩写 – 全称 – 学科类别对照表

对下载的纳米领域与纳米技术领域的文献的引文进行整理,得到 WoS 纳米科学与纳米技术领域文献引文信息(见表 2),其中 ID 表示文献的编号。由于下载的引文信息仅有引文所属的期刊,没有期刊所属的学科类别,所以需构建期刊 – 学科类别对照表。WoS 员工基于期刊标题及其引用图谱一系列指标为 WoS 数据库中所收录的每个期刊分配其对应的学科类别(Web of Science Category, WC),WoS 的《期刊引用报告》(*Journal Citation Reports*,JCR)网站上提供了每个期刊所属的学科类别信息[20]。利用 Python 下载 JCR 所提供的期刊以及其所属的学科类别信息构建期刊 – 学科类别对照表。由于从 JCR 中期刊信息使用的是期刊的全名表示,而从 WoS 中下载的纳米科学与纳米技术领域文献的题录信息中参考文献所属期刊使用该期刊的缩写名称表示,所以还需构建期刊全名 – 缩写名对照表。ESI Journal List 是基本科学指标数据库(Essential Science Indicators, ESI)发布的包含 WoS 数据库中期刊和期刊缩写信息的列表[21]。将从 JCR 数据库中获得的期刊全称 – 学科类别对照表与 ESI Journal List 的期刊全称 – 缩写对照表进行关联,得到构建期刊全称 – 期刊缩写 – 所属学科类别对照表,其中包含 WoS 收录的 11 375 个期刊,最终得到 17 961 个期刊 – 学科类别映射结果,即平均一个期刊对应 1.58 个学科类别。期刊缩写 – 全称 – 所属学科类别对照表如表 3 所示。

表 2 WoS 纳米科学与纳米技术领域文献引文信息(部分)

ID	参考文献	参考文献作者	参考文献年份	期刊	DOI
1	Alonso MI, 2005, J APPL PHYS, V98	M. I. Alonso	2005	*J APPL PHYS*	DOI 10.1063/1.2006229
1	Barry SD, 2011, CHEM MATER, V23, P5096	S. D. Barry	2011	*CHEM MATER*	DOI 10.1021/cm202761k
1	Benchamekh R, 2014, PHYS REV B, V89	R. Benchamekh	2014	*PHYS REV B*	DOI 10.1103/PhysRevB.89.035307
1	Bostedt C, 2004, APPL PHYS LETT, V84, P4056	C. Bostedt	2004	*APPL PHYS LETT*	DOI 10.1063/1.1751616
1	Boykin BT, 2005, PHYS REV B, V71	B. T. Boykin	2005	*PHYS REV B*	
1	Boykin TB, 2007, J PHYS-CONDENS MAT, V19	T. B. Boykin	2007	*J PHYS-CONDENS MAT*	DOI 10.1088/0953 – 8984/19/3/036203
1	Bricogne G, 2006, INT TABLES CRYSTALLO, VB	G. Bricogne	2006	*INT TABLES CRYSTALLO*	
1	Buljan M, 2010, PHYS REV B, V82	M. Buljan	2010	*PHYS REV B*	DOI 10.1103/PhysRevB.82.235407
1	Bulutay C, 2007, PHYS REV B, V76	C. Bulutay	2007	*PHYS REV B*	DOI 10.1103/PhysRevB.76.205321
…	…	…	…	…	…
40491	Zhou Y, 2009, ADV FUNCT MATER, V19, P3293	Y. Zhou	2009	*ADV FUNCT MATER*	DOI 10.1002/adfm.200900702

表 3 期刊全称 – 缩写 – 学科类别对照表(部分)

期刊缩写	期刊全称	学科类别
2D MATER	*2D Materials*	MATERIALS SCIENCE, MULTIDISCIPLINARY
3 BIOTECH	*3 Biotech*	BIOTECHNOLOGY & APPLIED MICROBIOLOGY
4OR-Q J OPER RES	*4OR-A Quarterly Journal of Operations Research*	OPERATIONS RESEARCH & MANAGEMENT SCIENCE
AAPG BULL	*AAPG BULLETIN*	GEOSCIENCES, MULTIDISCIPLINARY
AAPS J	*AAPS Journal*	PHARMACOLOGY & PHARMACY
AAPS PHARMSCITECH	*AAPS PHARMSCITECH*	PHARMACOLOGY & PHARMACY
AATCC J RES	*AATCC Journal of Research*	MATERIALS SCIENCE, TEXTILES
AATCC REV	*AATCC REVIEW*	CHEMISTRY, APPLIED
AATCC REV	*AATCC REVIEW*	ENGINEERING, CHEMICAL
AATCC REV	*AATCC REVIEW*	MATERIALS SCIENCE, TEXTILES

4.3 参考文献的学科映射

对下载的文献题录信息进行数据清洗,删除摘要、关键词、参考文献等信息不全的文献,最终得到40 491 篇文献共 1 649 353 条参考文献记录,平均一篇文献引

韩正琪，刘小平，徐涵. 基于 Rao-Stirling 指数的学科交叉文献发现——以纳米科学与纳米技术为例[J]. 图书情报工作，2018,62(1):125－131.

用 40.73 篇参考文献。基于 MySQL 数据库软件,利用期刊缩写－全称－所属学科类别对照表,将所下载的每篇文献的参考文献根据其所属期刊映射至一个或多个 WoS 学科类别中。最终将 1 461 341 条参考文献记录映射至相应学科类别,成功匹配的参考文献占总参考文献的 88.60%,平均一篇参考文献所属期刊对应 2.24个学科类别。

5 纳米科学与纳米技术领域学科交叉文献发现

5.1 构建文献参考文献的学科类别矩阵

利用 Python 编程,将文献参考文献所属学科类别转换为矩阵的形式表示(见表 4),其中,每一行代表一篇文献的参考文献的学科分布,每一列表示每篇文献属于该学科类别参考文献数目。为了使文献学科交叉程度测度结果更加可信,需剔除参考文献所属学科类别总数小于 3 的参考文献[22]。最终得到 38 598 篇待测度的纳米科学与纳米技术领域的文献参考文献的学科类别矩阵,占所下载文献总数的 95.32%。

5.2 纳米科学与纳米科技领域的学科交叉文献发现

在 Rao-Stirling 公式中,dij 为学科类别 i 和学科类别 j 之间的距离。为保证研究结果的可靠性和可对比性,本研究采用基于 D. Chavarro 等[22]所提供的 WoS 学科类别相似矩阵,即按照 I. Rafols 等[23]提出的方法,基于 2007 年 JCR 光盘版中收录的 60 947 519 篇文献,构建其所属学科类别的共被引矩阵,计算学科类别间的余弦相似度,最终得到学科相似度矩阵。基于该相似度矩阵编写 R 程序,分别测度每篇文献的种类(Variety)、均衡度(Balance)、差异性(Disparity)以及 Rao-Stirling 多样性指数。

表 4 文献参考文献学科类别矩阵(部分)

ID \ 学科类别	应用物理	凝聚态物理	纳米科学与纳米技术	物理化学	光学	工程、电气与电子	物理,原子、分子和化学	声学	……	动物学
1	24	20	12	10	4	2	2	0	……	0
2	19	16	25	22	3	1	4	0	……	0
3	4	3	6	17	0	0	3	0	……	0
4	6	11	6	14	0	1	5	0	……	0
5	7	12	2	5	1	0	1	0	……	0
6	6	15	12	17	0	0	4	0	……	0
7	6	7	8	12	2	0	5	0	……	0
8	32	14	19	18	4	0	1	0	……	0
9	8	4	9	9	0	0	0	0	……	0
10	7	0	7	6	0	1	0	0	……	0

表 5 所表示的是按照文献 Rao-Stirling 指数从大到小的部分测度结果,Rao-Stirling 指数越大,则该文献学科交叉程度越高。图 1 为纳米科学与纳米技术领域文献 Rao-Striling 指数的频率分布直方图。可以发现,纳米科学与纳米技术领域文献的 Rao-Stirling 指数大多集中于 0.2－0.5 之间。结合直方图并查阅相关文献可以将文献学科交叉分为三个水平:大部分文献的的 Rao-Stirling 指数在 0.2－0.5 之间,属于中度学科交叉文献;大于 0.5 的文献属于高度学科交叉文献;小于 0.2的文献属于低度学科交叉文献。可以根据研究需要设定不同的阈值,筛选领域学科交叉文献用于进一步的学科交叉主题识别。

表 5 纳米科学与纳米技术领域文献学科交叉测度结果

ID	标题	Rao-Stirling	种类	均衡度	差异性
29215	QUALITATIVE DATA ANALYSIS IN PRODUCT DEVELOPMENT - AN EXPLORATION OF CLOSED LOOP THINKING IN PRODUCT REQUIREMENTS	0.81	15	0.99	0.87
8050	Settling into the midstream? Lessons for governance from the decade of nanotechnology	0.80	19	0.96	0.85
40128	Mean Particle Diameters. Part VIII. Computer Program to Decompose Mixtures of (Truncated) Lognormal Particle Size Distributions Using Differential Evolution to Generate Starting Values for Nonlinear Least Squares	0.78	25	0.92	0.84
28877	A REVIEW ON THE NON-INVASIVE EVALUATION OF SKELETAL MUSCLE OXYGENATION	0.77	21	0.95	0.80
13144	Study of the Airflow Patterns and of the Characteristics of Bio-Aerosol Nanoparticle Deposition in Human Upper Respiratory Tracts Based on Computed Tomography Scanning Reconstruction	0.77	28	0.93	0.81
25077	Techniques and Applications of Emotion Recognition in Speech	0.77	13	0.96	0.83
39339	Detection of airborne bacteria with disposable bio-precipitator and NanoGene assay	0.77	26	0.93	0.79
21306	An Electrochemical Microbubble-Based MEMS Pressure Sensor	0.77	26	0.92	0.82
20794	Mental Fatigue Monitoring Using a Wearable Transparent Eye Detection System	0.76	17	0.95	0.83
…	…	…	…	…	…
40336	Stress relaxation in pulsed DC electromigration measurements	0.00	1	0.00	0.00

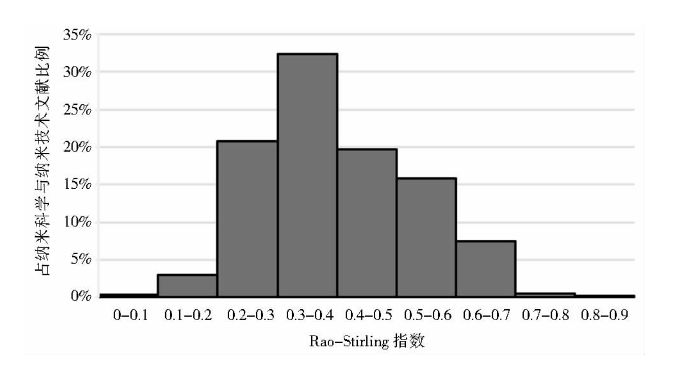


图 1 纳米科学与技术领域文献 Rao-Stirling 指数频率直方图

6 讨论与结论

学科交叉文献发现是进行学科交叉研究的重要前提,从海量的文献中快速、精准地发现领域相关交叉文献有助于研究人员快速地把握领域学科交叉动态,识别领域学科交叉研究热点与前沿。本研究基于引文分析,将 Rao-Stirling 指数用于学科交叉的文献发现研究,并以纳米科学与纳米技术领域为例,验证方法的可行性。另外,本研究所提出的学科交叉的文献发现方法以及构建的期刊缩写 – 全称 – 学科类别对照表和基于 Python 和 R 编写的参考文献学科映射和学科交叉测度指标的计算程序同样适用于其他领域的学科交叉文献发现研究,所得到的纳米科学与纳米技术领域的高度学科交叉文献也可用于后续的纳米科学与纳米技术领域的学科交叉主题识别研究,为后续的研究奠定基础。

本研究所提出的基于 Rao-Stirling 指数的学科交叉文献发现方法通过测度文献参考文献的学科分布情况相比前文所综述的其他方法具有以下优点:①该方法可以发现与某领域相关的所有学科交叉文献,且算法复杂度较低,有利于大数据集的学科交叉文献发现研究的实现。②基于文献引文所属学科类别对文献的学科交叉程度进行测度,可以从源头上发现研究者或者研究团队进行某一研究时所受到知识启发都来自哪些其他学科,从而能够更好的把握学科交叉的发展路线和未来发展方向。③该方法不仅可以发现学科交叉文献,还可以对学科交叉文献的学科交叉程度进行定量的测度,从而筛选得到高度学科交叉文献。一方面,目前研究发现高度学科交叉文献通常伴随着高被引量,也就是说高度的学科交叉文献通常代表着该领域当前研究热点和前沿[24];另一方面,由于目前文献或多或少都具有学科交叉特性,仅筛选出某领域高度学科交叉文献作为后续领域学科交叉主题识别研究对象可以减少数据量,降低算法复杂度。

本研究所提出的基于 Rao-Stirling 指数的学科交叉文献发现方法在整个研究过程中仍然存在一些问题有待进一步解决:一是本研究基于 WoS 的学科类别来判断文献参考文献所属学科,不同的学科分类体系测度文献学科交叉程度得到的结果也不完全相同;二是本研究所提出的基于引文分析的领域学科交叉文献发现是否能够准确的从认识论或者认知意义上识别学科交叉文献还有待于进一步验证;三是学科交叉文献阈值的确定,尚未有标准的阈值的划定某一文献是否为学科交叉文献或者高度学科交叉文献,虽然通过抽取 Rao-Stirling 指数大于等于 0.5 的文献进行通读后发现其具有较高的学科交叉性,但相关结果须要领域专家结合领域相关知识进行进一步验证。

综上,在未来研究中,计划进行基于多种学科分类体系的学科交叉测度对比研究,并将纳米科学与纳米技术领域学科交叉文献发现结果经专家判读后用于后续领域学科交叉的主题识别研究,以期发现纳米科学与纳米技术领域的学科交叉点。此外,拟开展其他学科的学科交叉文献发现研究。

参考文献:

[1] QIN J, LANCASTER F W, ALLEN B. Types and levels of collaboration in interdisciplinary research in the sciences[J]. Journal of the American Society for Information Science, 1997,48(10):893 –916.

[2] 路甬祥. 学科交叉与交叉科学的意义[J]. 中国科学院院刊, 2005,20(1):58 –60.

[3] 惠森. 诺贝尔自然科学奖获奖成果中的学科交叉现象研究[D]. 郑州:郑州大学, 2014.

[4] RIP A, COURITIAL J P. Co-word maps of biotechnology: an example of cognitive scientometrics[J]. Scientometrics, 1984, 6(6):381 –400.

[5] 魏建香. 学科交叉知识发现及其可视化研究[D]. 南京:南京大学, 2010.

[6] 李长玲, 郭凤娇, 魏绪秋. 基于时序关键词的学科交叉研究主题分析——以情报学与计算机科学为例[J]. 情报资料工作, 2014(6):44 –48.

[7] 吴蕾, 孙巍. 学科交叉热点主题发现与演化分析方法研究——以动物资源与育种领域为例[J]. 数字图书馆论坛, 2015(12):15 –20.

[8] ALLAN A. A method for determining interdisciplinary activities within a university[J]. Library research, 1980,2(1):83 –94.

[9] TIJSSEN R J W. A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research[J]. Research policy, 1992,21(1):27 –44.

[10] SCHUMMER J. Multidisciplinarity, interdisciplinarity, and pat-

terns of research collaboration in nanoscience and nanotechnology [J]. Scientometrics, 2004,59(3):425 - 465.

[11] CHUBIN D E, PORTERA L, ROSSINIF A. "Citation classics" analysis: an approach to characterizing interdisciplinary research [J]. Journal of the American Society for Information Science, 1984,35(6):360 - 368.

[12] PORTER A L, CHUBIND E. An indicator of cross-disciplinary research[J]. Scientometrics, 1985,8(3/4):161 - 176.

[13] CHOI J M. Citation analysis of intra- and interdisciplinary communication patterns of anthropology in the U. S. A. [J]. Behavioral &social sciences librarian, 1988,6(3 - 4):65 - 84.

[14] 魏海燕，尹怀琼，刘莉. 基于引文分析的情报学与相关学科的研究[J]. 情报杂志，2010,29(2):38 - 43.

[15] 杨良斌，周秋菊，金碧辉. 基于文献计量的跨学科测度及实证研究[J]. 图书情报工作，2009,53(10):87 - 90.

[16] 郭凤娇，李长玲，王晓笛，等. 基于向量空间模型的学科交叉文献发现研究——以情报学与计算机学科为例[J]. 情报杂志，2014(3):171 - 175.

[17] MUGABUSHAKA A M, KYRIAKOU A, PAPAZOGLOU T. Bibliometric indicators of interdisciplinarity: the potential of the Leinster——Cobbold diversity indices to study disciplinary diversity [J]. Scientometrics, 2016, 107(2):593 - 607.

[18] STIRLING A. A general framework for analysing diversity in science, technology and society[J]. Journal of the Royal Society interface, 2007,4(15): 707 - 719.

[19] LEYDESODORFF L. isi. exe[EB/OL]. [2017 - 07 - 04]. http://www.leydesdorff.net/software/isi/index.htm.

[20] Incites Journal Citation Reports[EB/OL]. [2017 - 07 - 04]. https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action.

[21] Incites Essential Science Indicators help[EB/OL]. [2017 - 07 - 04]. http://ipscience-help.thomsonreuters.com/incitesLiveESI/ESIGroup/overviewESI/esiJournalsList.html.

[22] CHAVARRO D, TANG P, RAFOLS I. Interdisciplinarity and research on local issues: evidence from a developing country[J]. Research evaluation, 2014,23(3): 195 - 209.

[23] RAFOLS I, PORTER A L, LEYDESDORFF L. Science overlay maps: a new tool for research policy and library management[J]. Journal of the Association for Information Science & Technology, 2010,61(9):1871 - 1887.

[24] CHENS, ARSENAULT C, LARIVIèRE V. Are top-cited papers more interdisciplinary? [J]. Journal of informetrics, 2015,9(4): 1034 - 1046.

作者贡献说明：

韩正琪：数据处理与分析，论文撰写；

刘小平：确定选题方向；论文写作指导与修订；

徐涵：数据收集。

Interdisciplinary Literature Discovery Based on Rao-Stirling Diversity Indices: Case Studies in Nanoscience and Nanotechnology

Han Zhengqi[1,2] Liu Xiaoping[2] Xu Han[1,2]

[1] University of Chinese Academy of Sciences, Beijing 100049

[2] National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] Discovering the interdisciplinary literature is an important prerequisite for interdisciplinary research. Finding domain-related interdisciplinary publications from vast amounts of literature quickly and accurately will help researchers understand the new tendency of interdisciplinarity and identify the focal and hot themes of the field rapidly. This paper presents a novel approach to discover the interdisciplinary literature based on Rao-Stirling diversity indices. Focusing on the case study in the field of nanoscience & nanotechnology, we discussed the feasibility of the method. [**Method/process**] Based on the nanoscience and nanotechnology publications in the Web of Science, we mapped references to the Web of Science Category, constructed the subject distribution matrix and calculated the interdisciplinary measure indices of the documents based on Python and R to find the interdisciplinary literature. [**Result/conclusion**] The paper shows that the interdisciplinary literature discovery method based on Rao-Stirling diversity indices can measure the degree of interdisciplinarity in the publication level and discover the interdisciplinary publications in the field of nanoscience and nanotechnology. The method could be applied to other research fields.

Keywords: interdisciplinarity Rao-Stirling interdisciplinary literature nanoscience and nanotechnology